A posteriori uncertainty quantification using MCMC for wildland fire spread simulation

Frédéric Allaire[a,*], Vivien Mallet[a], Jean-Baptiste Filippi[b]

[a] *Institut national de recherche en informatique et en automatique (INRIA), 2 rue Simone Iff, Paris, France; Sorbonne Université, Laboratoire Jacques-Louis Lions, France.*
[b] *Centre national de la recherche scientifique (CNRS), Sciences pour l'Environnement – Unité Mixte de Recherche 6134, Università di Corsica, Campus Grossetti, Corte, France.*

Abstract

Simulation is used to predict wildland fire spread in real-time. Nevertheless, the large uncertainties in these simulations must be quantified in order to provide better information to fire managers. Ensemble forecasts are usually applied for this purpose, with an input parameter distribution that is defined based on expert knowledge.

We propose a novel approach to generate calibrated ensembles whose input distribution is defined by a posterior PDF with a pseudo-likelihood function that involves the Wasserstein distance between simulated and observed burned surfaces of several fire cases. Due to the high dimension and the computational requirements of the pseudo-likelihood function, a Gaussian process emulator is built to obtain a sample of the calibrated input distribution with a MCMC algorithm in about one day of computation on 8 computing cores.

The calibrated ensembles lead to better overall accuracy than the uncalibrated ensembles. The a posteriori probability distribution of the inputs favors lower values of rate of spread and lower uncertainty on wind direction. This strongly limits overprediction, while keeping the ability of the ensemble to cover the observed burned area.

*Corresponding author
Email address: `frederic.allaire@inria.fr` (Frédéric Allaire)

Preprint submitted to Applied Mathematical Modelling

Keywords: uncertainty quantification, Metropolis Hastings, Wasserstein distance, Gaussian process

1. Introduction

Modeling wildland fire spread is a challenging task due to the high nonlinearity of the phenomenon and the significant uncertainties in the modeling process. Several models have been developed to describe the dynamics of wildland fire spread [1] with varying degrees of complexity leading to (semi-)physical and (semi-)empirical models. Physical models are too complex and the associated simulations are too time-expensive for the computation of a very large scale wildfire in real time. Meanwhile, 2D fire spread simulators [2] that describe the dynamics of the shape of the fire are faster and more suited to make predictions in an operational context. Such simulators typically make use of an empirical model where the rate of spread (ROS), i.e., the speed at which the flames advance, is expressed as a function of local environmental parameters (such as wind, slope, fuel moisture and vegetation properties).

Although the use of empirical models implies a drastic simplification of the physics of wildland fire spread, they are usually non-linear. For instance, most ROS models imply a power-type relation between ROS and wind speed [3].

High variability of the environmental conditions, difficulty of measurements, etc., lead to considerable uncertainty. Instead of relying on a single deterministic prediction, an alternative consists in generating a probabilistic prediction of fire spread in order to quantify this uncertainty. In wildland fire predictions, probabilistic methods mostly focus on the uncertainty of the inputs, which is propagated through the fire spread simulators (e.g., [4, 5, 6, 7, 8]). Other sources of uncertainty (notably, model errors) are usually unaccounted for.

Performance of fire spread simulators is typically assessed by comparing a simulated burned surface with its observed counterpart by the means of deterministic methods, for instance by computing indices that measure how much the two surfaces match (e.g. [9, 10])

or by analyzing the distance between vertices of the fire perimeter (e.g. [11, 12]). In the case of probabilistic predictions, several evaluation criteria we proposed in [13] to assess the performance of the prediction system.

Input uncertainty is typically quantified *a priori* via a probability distribution that is based on data measurements and expert knowledge. A way to improve the prior distribution is to compare the corresponding distribution of the model outputs with observations. The goal is to use the observational information to obtain a distribution *a posteriori* for both the input and the output. This procedure can be seen as a calibration of the input distribution and consists in solving a problem of *inverse uncertainty quantification* (e.g., [14]). This approach is relatively new in wildland fire modeling, although the Generalized Likelihood Uncertainty Estimation (GLUE) methodology was investigated in [15].

The goal of this study is to calibrate the probability distribution of the inputs of the model based on observed fires. The probabilistic predictions generated with the calibrated distributions should lead to better probabilistic scores than those generated with the prior distribution on the input variables. Two major difficulties are encountered: first, the large number of uncertain input variables; second, that the model is considered a "black box" whose output is a surface, so that we cannot easily write a formula for the likelihood, which is fundamental in a Bayesian approach. We propose a method inspired from the Bayesian framework that can circumvent these two difficulties by making use of a novel score for the comparison of surfaces relying on the Wasserstein distance for several observed fires. This leads to the definition of a calibrated distribution that can be sampled from via traditional Markov chain Monte Carlo (MCMC) algorithms. We use an emulator (also called surrogate model or metamodel) of the score to drastically decrease computational time, at the cost of relatively low approximation error, so that sufficient MCMC iterations can be performed in a reasonable amount of time.

This methodology follows the setting presented in a previous study [13] where the un-

certain inputs were identified. We try to obtain calibrated probabilistic predictions that are as relevant as possible based on the observation of 7 relatively big Corsican fires.

The theory behind the strategy for calibration of input uncertainty is described in Section 2 and the technical aspects of its application are presented in Section 3. In Section 4, we present the results obtained regarding the emulation, the calibration of the distributions and the evaluation of the resulting ensembles. The results are then discussed in Section 5.

2. A posteriori uncertainty quantification

In this section, we will first consider the observation of one burned surface. Application to observations of several fires will be introduced in subsection 2.3. We denote the observed burned surface as $\mathcal{S}_{obs}$. We can model fire spread with a numerical model $\mathcal{M}$, whose inputs may vary according to a vector $\boldsymbol{u}$ of d perturbations applied to reference inputs. The model is dynamical and may return burned surfaces at different times but we only focus on a surface that corresponds to the (estimated) observation time of $\mathcal{S}_{obs}$. The simulated burned surface being denoted as $\mathcal{S}_{\boldsymbol{u}}$, we have $\mathcal{S}_{\boldsymbol{u}} = \mathcal{M}(\boldsymbol{u})$. We can directly compare $\mathcal{S}_{\boldsymbol{u}}$ and $\mathcal{S}_{obs}$, but there is uncertainty on the input variables. This uncertainty is modeled by attributing a probability distribution to the perturbation vector, that can be seen as a random vector $\boldsymbol{U}$. Consequently, the output is also stochastic: $\mathcal{S}_{\boldsymbol{U}} = \mathcal{M}(\boldsymbol{U})$. Although $\mathcal{S}_{\boldsymbol{U}}$ is probabilistic and $\mathcal{S}_{obs}$ is deterministic, they can be compared by the means of probabilistic evaluation tools (see Appendix A for more details). We seek a distribution that is as suitable as possible for the random vector $\boldsymbol{U}$. We consider that the distribution of $\boldsymbol{U}$ is described by the probability density function (PDF) g and that we already have access to a *prior* density function f for $\boldsymbol{U}$. In this section, we propose a method to obtain g by making the best possible use of f, $\mathcal{S}_{obs}$ and $\mathcal{M}$.

2.1. Distribution based on Wasserstein distance

A classical choice for g would be the *posterior* density function $p(.|\mathcal{S}_{obs})$ that is obtained according to Bayes' rule:

$$p(\boldsymbol{u}|\mathcal{S}_{obs}) = \frac{\mathcal{L}(\mathcal{S}_{obs}|\boldsymbol{u})f(\boldsymbol{u})}{\int \mathcal{L}(\mathcal{S}_{obs}|\boldsymbol{u})f(\boldsymbol{u})d\boldsymbol{u}}, \tag{1}$$

where $\mathcal{L}(\mathcal{S}_{obs}|\boldsymbol{u})$ would be the likelihood of the observation $\mathcal{S}_{obs}$ knowing the perturbation vector $\boldsymbol{u}$. However, defining the likelihood requires to make an appropriate probabilistic hypothesis, where $\mathcal{S}_{obs}$ is a realization of a 2D stochastic process whose distribution depends on $\boldsymbol{u}$. Making such a hypothesis is not trivial, but a step in this direction would be to use $\mathcal{S}_{\boldsymbol{u}}$ rather than simply $\boldsymbol{u}$, and define a (conditional) probability distribution for $\mathcal{S}_{obs}$ based on $\mathcal{S}_{\boldsymbol{u}}$. A desirable property of such a probability distribution is that the most likely realizations of $\mathcal{S}_{obs}$ are the ones that are most similar to $\mathcal{S}_{\boldsymbol{u}}$. Also, similarity should take into account high correlation between two points in a 2D domain when they are close. For instance, if a given location has high probability of being burned, so should have its neighboring locations. Still, while defining a likelihood for a vector is feasible, this might not be the case for a random surface.

Therefore, we propose a calibrated distribution that is inspired from Bayes' rule, where the density function g can be written in the following form:

$$g_{E,\beta}(\boldsymbol{u}) = \frac{e^{-\beta E(\boldsymbol{u})}f(\boldsymbol{u})}{\int e^{-\beta E(\boldsymbol{u})}f(\boldsymbol{u})d\boldsymbol{u}}, \tag{2}$$

where $\beta > 0$ and E is a positive "energy" function that is equal to 0 when $\mathcal{S}_{\boldsymbol{u}} = \mathcal{S}_{obs}$ and increases with the dissimilarity between $\mathcal{S}_{\boldsymbol{u}}$ and $\mathcal{S}_{obs}$. Here, we have a pseudo-likelihood function that plays the role of $\mathcal{L}$ in equation (1). This calibrated family of functions is inspired from Gibbs measures, but is different because the exponential is multiplied by f,

the prior PDF. The higher the parameter β, the more weight is given to the pseudo-likelihood function. Also, when $\beta = 0$, the calibrated PDF is equal to the prior PDF.

Several scores to compare $\mathcal{S}_{obs}$ and $\mathcal{S}_{\boldsymbol{u}}$ exist and could be used directly or after minor modifications to make suitable choices for E. We decided to introduce a novel score that makes use of the Wasserstein distance, which is a metric between probability distributions. The square of the Wasserstein distance between the uniform probability distributions covering respectively $\mathcal{S}_{obs}$ and $\mathcal{S}_{\boldsymbol{u}}$ is our choice for $E(\boldsymbol{u})$ and can be roughly defined as follows:

$$E(\boldsymbol{u}) = \inf_{\psi} \Bigg\{ \int_{\mathcal{S}_{obs} \times \mathcal{S}_{\boldsymbol{u}}} ||x - y||_2^2 \ \psi(x,y) dx dy \ \Bigg| \ \int_{\mathcal{S}_{obs}} \psi(x,y) dx = \frac{\mathbb{1}(x \in \mathcal{S}_{\boldsymbol{u}})}{|\mathcal{S}_{\boldsymbol{u}}|}, \int_{\mathcal{S}_{\boldsymbol{u}}} \psi(x,y) dy = \frac{\mathbb{1}(y \in \mathcal{S}_{obs})}{|\mathcal{S}_{obs}|} \Bigg\}, \tag{3}$$

where $\mathbb{1}$ stands for the indicator function, $||.||_2$ is the Euclidean distance (here, in $\mathbb{R}^2$), and $|\mathcal{S}|$ is the surface area of $\mathcal{S}$. It can be thought of the minimum energy that is required to move the points contained in $\mathcal{S}_{obs}$ so as to transform the surface into $\mathcal{S}_{\boldsymbol{u}}$. Also, when both surfaces are the same, we have $E(\boldsymbol{u}) = 0$. More details regarding the definition and the numerical approximation of the Wasserstein distance are given in appendix Appendix B.

The denominator of $g_{E,\beta}(\boldsymbol{u})$ is a high-dimensional integral that is intractable but that does not depend on the perturbation vector $\boldsymbol{u}$. Hence for a given β, the PDF is known up to some constant factor. When a distribution is known up to a factor, the Metropolis-Hastings (MH) algorithm allows to draw samples from that distribution. In this paper, we will therefore employ the MH algorithm, and describe the a posteriori distribution $\boldsymbol{U}$ with a very large sample. Computing $\mathcal{S}_{\boldsymbol{u}}$ and the Wasserstein distance to obtain $E(\boldsymbol{u})$ can be done in a reasonable amount of time. Nonetheless, MH algorithm may require a lot of iterations ($\sim 10^5$) to obtain a sufficiently large sample, which would take too much time. To speed up the MH algorithm, we propose to use an emulator $\tilde{E}$ instead of computing E exactly.

The emulator $\tilde{E}(\boldsymbol{u})$ will be a good approximation of $E(\boldsymbol{u})$ and will be considerably faster. The general design of the emulator is explained in Section 2.2 and its application to an energy function specific to several fire cases is given in Section 2.3. To determine whether an appropriate sample is returned by the MH algorithm, we will use the multivariate diagnostic metric proposed by [16]. The details on this procedure are given in Section 2.4.

2.2. Emulation

The focus of this section is the approximation of a function $y : \boldsymbol{u} \in \mathcal{D} \subset \mathbb{R}^d \to \mathbb{R}$. This function can be the previous Wasserstein distance $E(\boldsymbol{u})$, or the extension to several fires presented in Section 2.3. Note that the following emulation approach is fairly general and can be applied to a wide range of functions.

2.2.1. Gaussian process modeling

The emulation method used in this study is Gaussian process (GP) modeling, also called kriging. In this context, $y(\boldsymbol{u})$ is seen as a realization of a Gaussian process $Y_{\boldsymbol{u}}$ indexed by $\boldsymbol{u}$. Its means that any random vector $[Y_{\boldsymbol{u}^1}, \ldots, Y_{\boldsymbol{u}^n}]^T$ with $n < \infty$ components follows a Gaussian multivariate distribution. We denote a the trend function of the process, i.e., $\mathbb{E}[Y_{\boldsymbol{u}}] = a(\boldsymbol{u})$. The centered process $Z_{\boldsymbol{u}} = Y_{\boldsymbol{u}} - a(\boldsymbol{u})$ is also Gaussian, with a covariance function of the form $\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{u}') = \sigma^2 \rho(\boldsymbol{u} - \boldsymbol{u}')$, where $\sigma^2 > 0$ and ρ is the correlation function between two input points $\boldsymbol{u}$ and $\boldsymbol{u}'$.

We have at our disposal a set of training data $\left(\boldsymbol{u}^i, y(\boldsymbol{u}^i)\right)_{i=1,\ldots,n}$. We denote $\boldsymbol{Y}^n = [Y_{\boldsymbol{u}^1}, ..., Y_{\boldsymbol{u}^n}]^T$ and $\boldsymbol{y}_n = [y(\boldsymbol{u}^1), ..., y(\boldsymbol{u}^n)]^T$. We define R_n as the correlation matrix on the inputs of the training data:

$$R_n = \left(\rho(\boldsymbol{u}^i - \boldsymbol{u}^j)\right)_{1 \geq i,j \geq n}, \tag{4}$$

and $\boldsymbol{a}_n = [a(\boldsymbol{u}^1), ..., a(\boldsymbol{u}^n)]^T$ as the vector of trends in the training data.

For a new point $\boldsymbol{u}^*$ (outside of the training sample or not), we define the correlation vector $\boldsymbol{r}^* = [r(\boldsymbol{u}^* - \boldsymbol{u}^1), ..., r(\boldsymbol{u}^* - \boldsymbol{u}^n)]^T$. Under the assumptions made on $Y_{\boldsymbol{u}}$, the joint probability distribution between $\boldsymbol{Y}^n$ and $Y_{\boldsymbol{u}^*}$ is Gaussian and so is the conditional distribution of $Y_{\boldsymbol{u}^*}$ knowing $\boldsymbol{Y}^n$. We have

$$Y_{\boldsymbol{u}^*}|\boldsymbol{Y}^n \sim \mathcal{N}\big(\mathbb{E}[Y_{\boldsymbol{u}^*}|\boldsymbol{Y}^n], \mathrm{Var}[Y_{\boldsymbol{u}^*}|\boldsymbol{Y}^n]\big), \tag{5}$$

where

$$\mathbb{E}[Y_{\boldsymbol{u}^*}|\boldsymbol{Y}^n] = a(\boldsymbol{u}^*) + \boldsymbol{r}^{*T} {R_n}^{-1}(\boldsymbol{y}_n - \boldsymbol{a}_n) \tag{6}$$

and

$$\mathrm{Var}[Y_{\boldsymbol{u}^*}|\boldsymbol{Y}^n] = \sigma^2(1 - \boldsymbol{r}^{*T} {R_n}^{-1}\boldsymbol{r}^*). \tag{7}$$

For any $\boldsymbol{u}^* \in \mathcal{D}$, we define an emulator $\tilde{y}$ of y as the mean of the conditional variable given by equation (6):

$$\tilde{y}(\boldsymbol{u}^*) = a(\boldsymbol{u}^*) + \boldsymbol{r}^{*T} {R_n}^{-1}(\boldsymbol{y}_n - \boldsymbol{a}_n). \tag{8}$$

In the present case, we choose a linear trend for $Y_{\boldsymbol{u}}$, i.e., $\mathbb{E}[Y_{\boldsymbol{u}}] = a(\boldsymbol{u}) = \alpha_0 + \boldsymbol{u}^T\boldsymbol{\alpha}$ where $\alpha_0 \in \mathbb{R}$ and $\boldsymbol{\alpha} \in \mathbb{R}^d$. We also choose the correlation function to be a product of one-dimensional Matérn 5/2 correlation functions, i.e.,

$$\forall \boldsymbol{u}, \boldsymbol{u}' \in \mathcal{D}, \rho(\boldsymbol{u} - \boldsymbol{u}') = \prod_{l=1}^{d} \left(1 + \frac{\sqrt{5}|u_l - u_l'|}{\theta_l} + \frac{5|u_l - u_l'|^2}{3{\theta_l}^2}\right) \exp\left(-\frac{\sqrt{5}|u_l - u_l'|}{\theta_l}\right), \tag{9}$$

where $\theta_1, ..., \theta_d > 0$.

For the sake of clarity, we previously presented simple kriging, where the coefficients a and the covariance hyperparameters are known. In this study, we use universal kriging where the trend is an unknown polynomial (for more information, see for instance [17]). In

practice, the $2d+2$ hyperparameters $\sigma^2, \alpha_0, \alpha_1, ..., \alpha_d, \theta_1, ..., \theta_d$ used to define the Gaussian process are unknown and can be estimated as the maximum likelihood estimators for the training dataset [18].

2.2.2. Design of experiments and performance metrics

The inputs of the training sample are obtained via a Latin hypersquare sample (LHS) with optimized discrepancy. As the GP emulator is interpolating from the points of the training sample, we also generate a complementary test sample to evaluate the predictive performance of the emulator far from the training points. The complementary sample is obtained with the algorithm for an optimal validation design described in [19]. It relies on a Halton sequence whose points are selected in order to keep a low discrepancy when both training and test samples are taken together. This procedure aims at selecting points that are located far from each other but also far from the points of the training sample, where we expect the approximation error to be higher.

Based on the test sample $\left(\boldsymbol{u}^i, y(\boldsymbol{u}^i)\right)_{i=1,...,n_{test}}$ several performance metrics can be used to evaluate the emulator $\tilde{y}$. In this study, we use the mean absolute error (MAE), defined as

$$MAE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} |\tilde{y}(\boldsymbol{u}^i) - y(\boldsymbol{u}^i)| \tag{10}$$

and the standardized mean square error (SMSE [20]) defined as:

$$SMSE = \frac{\sum_{i=1}^{n_{test}} \left(\tilde{y}(\boldsymbol{u}^i) - y(\boldsymbol{u}^i)\right)^2}{\sum_{i=1}^{n_{test}} \left(y(\boldsymbol{u}^i) - \bar{y}\right)^2}, \tag{11}$$

where $\bar{y} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} y(\boldsymbol{u}^i)$ is the sample mean of the emulated function based on the test sample. Both scores get closer to 0 as the error of the model gets lower. The SMSE can be seen as a mean squared error normalized by the variance of the function on the test sample. Note that a model that would always predict the mean of the training set would have an

SMSE approximately equal to 1.

2.3. Extension to several fire cases

If we have K fire cases, we can compute the energy functions $E_1, ..., E_K$ that correspond to each fire. An intuitive choice for the combined energy function is $E : \boldsymbol{u} \mapsto \sum_{k=1}^{K} E_k(\boldsymbol{u})$. However, a concerning issue is when the variations of $E_1(\boldsymbol{u})$ (for instance) are much higher than for the other fires. In this case, the variations of the pseudo-likelihood will be mostly determined by those of $E_1(\boldsymbol{u})$, and the calibrated distribution will be mostly representative of the information from the first fire at the expense of the other observations.

To circumvent this issue, we propose to weigh each fire depending on the values taken by $E_k(\boldsymbol{u})$. We define the energy function as the weighted sum of squared Wasserstein distances:

$$E(\boldsymbol{u}) = \sum_{k=1}^{K} w_k E_k(\boldsymbol{u}), \tag{12}$$

where the weights are defined using all points from the training dataset:

$$w_k = \frac{1}{\sum_{i=1}^{n} E_k(\boldsymbol{u}^i)}. \tag{13}$$

It is possible to emulate $E(\boldsymbol{u})$ directly but while the function is positive, emulation by GP does not guarantee positivity outside of the training sample. We choose instead to emulate $L(\boldsymbol{u}) = \log E(\boldsymbol{u})$ by the GP procedure described in Section 2.2, leading to the emulator $\tilde{L}(\boldsymbol{u})$. To emulate $E(\boldsymbol{u})$, we take the exponential $\tilde{E}(\boldsymbol{u}) = \exp \tilde{L}(\boldsymbol{u})$, which ensures positivity. The GP emulation is implemented in the R-package *DiceKriging* [17].

2.4. Sampling from the calibrated distribution

In this section, we present the procedure that we used to obtain a sample following a PDF of the form $g_{\tilde{E},\beta}$ described in equation (2). To run the algorithm in reasonable computational

time, we use the emulator $\tilde{E}$ in place of the energy function E as explained in section 2.3, assuming that the target distribution of MH is close enough to the desired distribution whose PDF is $g_{E,\beta}$.

Algorithm 1 Metropolis-Hastings algorithm applied to $g_{\tilde{E},\beta}$ (several chains)

Define m, n, and an instrumental distribution of PDF $q : \boldsymbol{u} \mapsto q(\boldsymbol{u}|\boldsymbol{v})$
for $j = 1, ..., m$ **do**
  Choose a starting point $\boldsymbol{u}_{1,j}$
  for $i = 2, ..., n$ **do**
    Sample a candidate $\boldsymbol{u}_{c,j} \sim q(.|\boldsymbol{u}_{i-1,j})$
    Compute the ratio

$$\tau = \frac{g_{\tilde{E},\beta}(\boldsymbol{u}_{c,j})\ q(\boldsymbol{u}_{i-1,j}|\boldsymbol{u}_{c,j})}{g_{\tilde{E},\beta}(\boldsymbol{u}_{i-1,j})\ q(\boldsymbol{u}_{c,j}|\boldsymbol{u}_{i-1,j})} = \frac{e^{-\beta\tilde{E}(\boldsymbol{u}_{c,j})}\ f(\boldsymbol{u}_{c,j})\ q(\boldsymbol{u}_{i-1,j}|\boldsymbol{u}_{c,j})}{e^{-\beta\tilde{E}(\boldsymbol{u}_{i-1,j})}\ f(\boldsymbol{u}_{i-1,j})\ q(\boldsymbol{u}_{c,j}|\boldsymbol{u}_{i-1,j})} \tag{14}$$

    if $\tau \geq 1$ **then**
      [Accept the candidate]
      $\boldsymbol{u}_{i,j} \leftarrow \boldsymbol{u}_{c,j}$
    else
      [Accept the candidate with probability τ]
      Sample $p \sim \mathcal{U}(0, 1)$
      if $p \leq \tau$ **then**
        $\boldsymbol{u}_{i,j} \leftarrow \boldsymbol{u}_{c,j}$
      else
        $\boldsymbol{u}_{i,j} \leftarrow \boldsymbol{u}_{i-1,j}$
      end if
    end if
  end for
end for
return $(\boldsymbol{u}_{1,j}, ..., \boldsymbol{u}_{n,j})_{j=1,...,m}$

Since we use the convergence diagnosis for MCMC algorithms introduced by Brooks and Gelman [16], Algorithm 1 presents a version of MH with several chains. It is recommended to choose the starting points $\boldsymbol{u}_{1,1}, ..., \boldsymbol{u}_{1,m}$ quite far from each other. The loop on the m chains can be parallelized easily. Based on the chains returned by the MH algorithm, we compute the between-sequence covariance matrix B/n (of size d) and the within sequence

covariance matrix W as follows:

$$B/n = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\boldsymbol{u}}_j - \bar{\boldsymbol{u}})(\bar{\boldsymbol{u}}_j - \bar{\boldsymbol{u}})^T, \tag{15}$$

$$W = \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{i=1}^{n}(\boldsymbol{u}_{i,j} - \bar{\boldsymbol{u}}_j)(\boldsymbol{u}_{i,j} - \bar{\boldsymbol{u}}_j)^T, \tag{16}$$

where $\bar{\boldsymbol{u}}_j = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{u}_{i,j}$ is the sample mean of the j-th chain, and $\bar{\boldsymbol{u}} = \frac{1}{m}\sum_{j=1}\bar{\boldsymbol{u}}_j$ is the sample mean over all chains. The metric used for analyzing convergence is

$$\hat{R}^d = \frac{n-1}{n} + \left(\frac{m+1}{m}\right)\lambda_1, \tag{17}$$

where λ_1 is the largest eigenvalue of the symmetric, positive definite matrix $W^{-1}B/n$. At convergence, $\hat{R}^d$ tends to 1, and following the recommendations of Gelman and Brooks [16], we consider that if $\hat{R}^d < 1.1$ when computed on the second half of the chains, a sufficient number of MH iterations has been performed. Furthermore, we choose to consider that the set comprising the second half of all m chains constitutes a representative sample of our target distribution when $\hat{R}^d < 1.1$.

3. Application to wildland fire propagation

3.1. Fire spread simulation

In this study, we use the fire propagation solver ForeFire [21]. ForeFire uses a front-tracking technique to model the propagation of the fire front, i.e., the interface between the burned surface and the rest of the simulation domain (not burned). The fire front is discretized by the means of Lagrangian markers linked by a dynamic mesh. Each marker is advanced according to the surface geometry and the rate of spread (ROS). Contrary to discrete time simulation methods, ForeFire relies on a discrete event specification. Each

marker is therefore advanced according to a given spatial increment and the time at which the marker will reach its next position is deduced from its propagation speed, making the simulation method asynchronous. Advancing a marker in time is considered as an event. Other events may lead to a new calculation of the future location and time advance of a marker, such as topology checks that determine whether the markers describe a proper burned surface and reshape the fire front when it is not the case.

In this study, the rate of spread is computed according to the empirical model of Rothermel [22], widely used in wildland fire simulation, and that contains numerous parameters already fitted and fixed through an analysis of a large set of laboratory experiments. The variables of this model are expressed in customary US units. The input variables are m_c, the fuel moisture content of dead fuel, S_v the surface-volume ratio, ΔH, the heat content, σ_f, the fuel load, M_χ, the moisture of extinction, ρ_p, the particle density, h, the fuel bed depth (denoted as fuel height in the following section), α, the slope, $\boldsymbol{n}$ the normal vector oriented in the direction of fire spread, and $\boldsymbol{W}$ the wind speed vector (predicted by the meteorological model). First of all, the ROS is given by equation (18)

$$ROS = \frac{I_r\ \xi\ (1 + \phi_V + \phi_P)}{\rho_d\ \epsilon\ Q_{ig}}. \tag{18}$$

The equation for heat of preignition reads

$$Q_{ig} = 250 + 1116\ m_c, \tag{19}$$

and that for the effective heating number reads

$$\epsilon = \exp(-138/S_v). \tag{20}$$

The reaction intensity is expressed as a product:

$$I_r = \Gamma' \; W_n \; \Delta H \; \eta_m \; \eta_s. \tag{21}$$

We assume that η_s, the mineral damping coefficient, is equal to 1. We also assume that the fuel mineral content is negligible, leading to $W_n = \sigma_f$. The moisture damping is expressed as a polynomial:

$$\eta_m = 1 - 2.59\frac{m_c}{M_\chi} + 5.11\left(\frac{m_c}{M_\chi}\right)^2 - 3.52\left(\frac{m_c}{M_\chi}\right)^3. \tag{22}$$

The equation for potential reaction velocity reads

$$\Gamma' = \frac{{S_v}^{1.5}}{495 + 0.0594{S_v}^{1.5}}\left(\frac{\beta}{\beta_{op}}\right)^A \exp\left(A\Big(1 - \frac{\beta}{\beta_{op}}\Big)\right), \tag{23}$$

with

$$A = (4.774{S_v}^{0.1} - 7.27)^{-1}, \tag{24}$$

where the packing ratio is given by

$$\beta = \frac{\rho_b}{\rho_p} = \frac{\sigma_f}{h\rho_p} \tag{25}$$

(note that the bulk density is given by $\rho_b = \frac{\sigma_f}{h}$), and the optimal packing ratio is expressed as

$$\beta_{op} = 3.348{S_v}^{-0.8189}. \tag{26}$$

Coming back to the numerator of equation (18), we have the following relationship for the

ratio of the no-wind propagation flux by the reaction intensity:

$$\xi = \frac{\exp\left((0.792 + 0.681{S_v}^{0.5})(\beta + 0.1)\right)}{192 + 0.259S_v} \tag{27}$$

Finally, we have the slope factor

$$\phi_P = 5.275\beta^{-0.3}\alpha^2 \tag{28}$$

and the wind speed factor

$$\phi_V = 7.47\exp(-0.133{S_v}^{0.55})\left(\frac{\beta}{\beta_{op}}\right)^{-0.715\exp(-0.000359S_v)} {W_S}^{0.0256{S_v}^{0.54}}, \tag{29}$$

where the "effective" wind speed, W_S, is computed via $\boldsymbol{W}$ and $\boldsymbol{n}$. To account for the fact that the wind speed at mid-height of the flame is usually lower than that of the prediction, we apply a 0.4 factor to wind speed. We also use the wind speed limit function proposed in [23], which leads to:

$$W_S = \min(0.4\boldsymbol{W}.\boldsymbol{n},\ 96.81{I_r}^{1/3}) \tag{30}$$

The scheme used to advance the markers of the fire front is based on a first-order approximation. Considering a marker that is located at x_i at time t_i, with its normal to the front denoted as $\boldsymbol{n}_i$ (oriented toward the unburned area), its next location is determined by

$$x_{i+1} = x_i + \delta l\ \boldsymbol{n}_i, \tag{31}$$

and the advance in time depends on ROS_i, the rate of spread computed with the values of the environmental inputs at location x_i and time t_i, as follows:

$$t_{i+1} = t_i + \frac{\delta l}{ROS_i}. \tag{32}$$

3.2. Prior uncertainty on input data

Input	Unit	Perturbation	Distribution	Notes
Wind direction	°	Additive	$\mathcal{N}(0, 60^2)$	Truncated to $[-180, 180]$
Wind speed norm	$\mathrm{m\,s^{-1}}$	Multiplicative	$\mathcal{LN}(0, (0.5 \log 3)^2)$	Truncated to $[1/3, 3]$
Dead fuel moisture		Multiplicative	$\mathcal{U}(0.4, 1.6)$	
Heat of combustion	$\mathrm{MJ\,kg^{-1}}$	Additive	$\mathcal{U}(-5, 5)$	
Particle density	$\mathrm{kg\,m^{-3}}$	Additive	$\mathcal{U}(-300, 300)$	
Fuel height	m	Multiplicative, individual	$\mathcal{U}(0.4, 1.6)$	
Fuel load	$\mathrm{kg\,m^{-2}}$	Multiplicative, individual	$\mathcal{U}(0.4, 1.6)$	
Surface-volume ratio	$\mathrm{m^{-1}}$	Multiplicative, individual	$\mathcal{U}(0.4, 1.6)$	
Direction from ignition point	°	Additive	$\mathcal{U}(-180, 180)$	
Distance to ignition point	m	Additive	$\mathcal{U}(0, 1) \times \Delta_{max}$	$\Delta_{max} \in \{100, 500, 1000\}$
Time of fire start	min	Additive	$\mathcal{U}(-1, 1) \times \Delta_{max}$	$\Delta_{max} \in \{10, 15, 30, 60\}$
Time of fire end	min	Additive	$\mathcal{U}(-1, 1) \times \Delta_{max}$	$\Delta_{max} \in \{10, 60, 120, 180, 240\}$

Table 1: Prior probability distribution of the perturbations on the simulation inputs.
For the first two inputs, the distribution is a truncated (log-)normal. For each of the last three inputs, we use a reduced variable in [0,1] or [-1, 1] in the calibration procedure, and multiply it by Δ_{max}, which depends on the fire case, when we run the simulations.

The description of the perturbation variables is presented in [13], where the distributions are mostly truncated normal. In the present study, we choose a wider support for some of the distributions and/or substitute them for uniform distributions. Another difference compared to the previous study is that we do not take the fuel type transitions in the calibration procedure, which means that no transition is applied when running the simulations to compute the emulator, and that transitions are not part of the MH algorithm. We still use them in the generation of the ensembles, by generating the transitions independently from the scalar perturbations.

The marginals of the prior distribution are presented in Table 1. We assume the inputs are independent. Note that in our simulations we consider up to 13 burnable fuel types that are linked to the Corine Land Cover classification [24]. When we indicate that the perturbation

is "individual", it means that we sample one perturbation coefficient for each fuel type. For one simulation, we generate $d = 48$ perturbation coefficients. For wind direction and wind speed norm, the distributions are truncated. For these two variables, the distribution specified in the "Distribution" column corresponds to the distribution before truncation is applied. The main reason behind the choice of distributions with finite support is to avoid sampling extreme values that may be unrealistic or that could lead to unphysical parameter values (e.g., negative fuel load).

Uncertainty on the location of the ignition point is specified by the perturbation parameters "direction from ignition point" and "distance to ignition point". To sample a perturbed ignition point, we first sample a direction and select the new ignition point at an independently sampled distance from the reference ignition point in this direction. The maximum distance Δ_{max} depends on the fire case: the perturbed ignition point may therefore be sampled within a radius ranging from 100 m to 1 km around the reference. Similarly, the maximum perturbation Δ_{max} for time of fire start and time of fire end depends on the fire case. This varying uncertainty is due to the information available regarding each fire. For the last three variables, Δ_{max} is specific to each fire case. However, for the calibration, there is only one "reduced" variable for each of these three inputs, whose support is either $[0, 1]$ or $[-1, 1]$. The actual perturbation used to run the fire spread simulations for a given fire case is obtained from the reduced variable after multiplication by Δ_{max}.

3.3. Application to seven Corsican wildland fires

The emulation and calibration procedure is applied to $K = 7$ fires that occurred in Corsica in 2017-2018 and that are presented in [13]. We also use the ensembles obtained in this previous study as reference for comparison and refer to them as "reference ensembles" hereafter.

We choose to build an emulator with a training sample of size 4000, and evaluate its

performance with a test sample of size 2000.

We then apply the MH algorithm to several distributions with different values of β ranging in $\{\frac{1}{10}, \frac{1}{7}, \frac{1}{4}, \frac{1}{2}, 1, 2\}$. We choose $n = 150000$ iterations and $m = 8$ chains. The instrumental distribution described by $q(\boldsymbol{u}|\boldsymbol{v})$ is a product of independent univariate truncated normal distributions. Before truncation, the k-th normal distribution is centered on the k-th component of $\boldsymbol{v}$ and has a standard deviation equal to a twentieth of the width of the perturbation range. The distribution is then truncated to the perturbation range. For the perturbation of wind speed norm, it is the logarithm of the perturbation that follows a truncated normal distribution.

By taking the latter half of the chains obtained with the MH algorithm, this leads to samples of size $m \times n/2 = 600000$ for each value of β. Based on these empirical distributions, we generate ensembles of wildland fire simulations for the seven fire cases, that we refer to as "calibrated ensembles" hereafter. The size of a calibrated ensemble ranges between 2000 and 10000. The evaluation domain is the same as for the reference ensembles, but contrary to the previous study, we do not apply any computational time limit to the simulations. The ensemble generation procedure is also carried out based on the prior distributions, which leads to "prior ensembles".

We evaluate the ensembles by following the methodology presented in [13]. We mainly focus on the accuracy of the ensembles by the means of the Brier skill score and a global version to summarize the performance of the ensembles on the seven fires. Bootstrap resampling is carried out to obtain 95% confidence intervals. More details can be found in appendix Appendix A. To summarize the information given by the other evaluation tools on several fires (rank histogram, reliability and sharpness diagrams), the contribution of each fire case is weighted by the size of the evaluation domain before summing the contributions of the seven fires (otherwise, fires with the largest simulation domains would have the most influence). For the rank histogram, we normalize the values of the rank because of the

varying ensemble size.

4. Results

4.1. Emulation

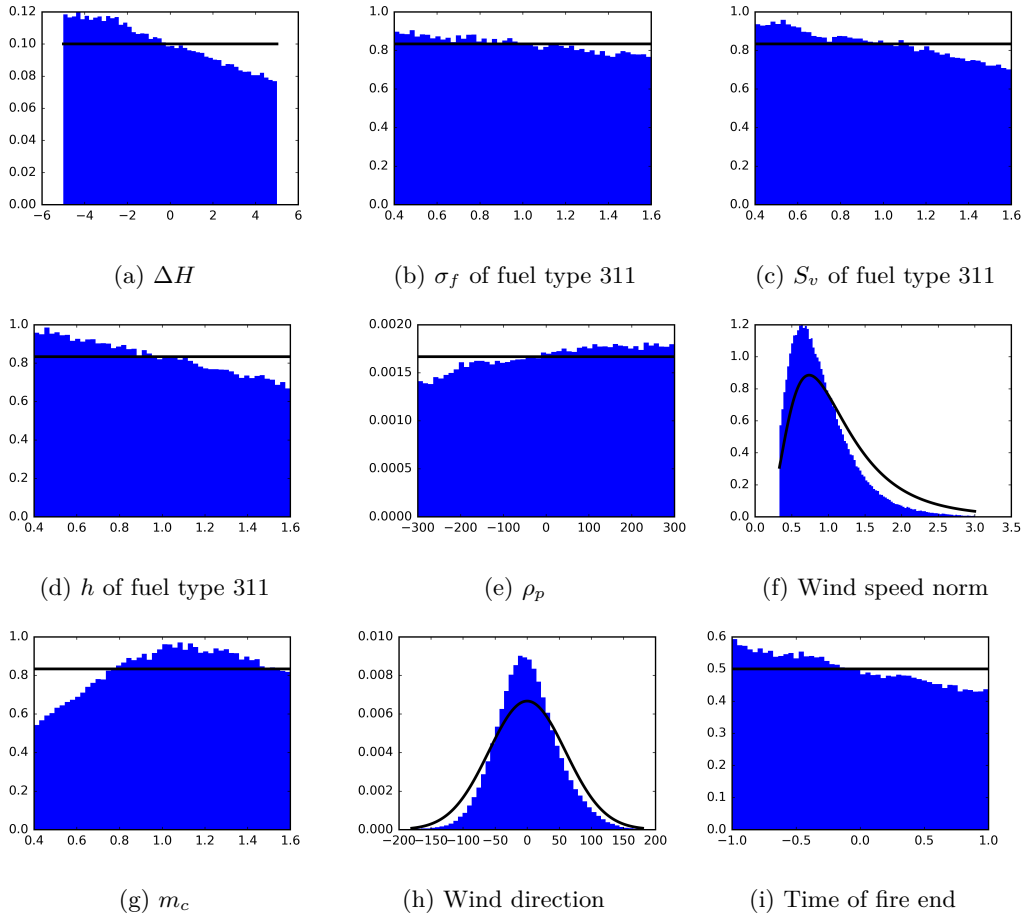
The emulator shows good predictive performance with $MAE = 0.73$ and $SMSE = 4.7\%$. The use of the logarithm allowed slightly better performance than with direct emulation of the energy function ($MAE = 0.97$ and $SMSE = 6.2\%$ without using the logarithm). In the latter case, negative values were obtained at 4 points of the test sample, while the use of the logarithm ensured the prediction of positive values. The emulated energy function is computed in approximately 0.6 s, therefore the 150000 iterations of the MH algorithm can be performed in a bit more than a day.

4.2. Calibrated distribution

For all values of β, our convergence diagnosis is positive with $\hat{R}^p$ ranging between 1.035 and 1.045 based on the last 75000 values of the chains. The proportion of accepted values ranges between 65% and 71% and decreases with β.

The correlation between different input variables of the calibrated distributions is low, with a maximum absolute value of 0.1046 (for $\beta = 2$). Also, most marginals show very little difference between the prior and the calibrated distribution. Although increasing β leads to more significant difference from the prior distribution, some variables seem to be unaffected by the calibration, namely most individual perturbations on fuel parameters (height, load and surface-volume ratio) and perturbation on the ignition point and the time of fire start.

Histograms of some marginals of the calibrated distribution are represented for $\beta = 1/2$ in Figure 1. Only the distributions that lead to the most significant change compared to the prior distribution are represented, namely perturbations of ΔH, ρ_p, the three individual fuel parameters (σ_f, S_v, h) specific to fuel type 311 (Broad-leaved forest), wind speed norm and



(a) ΔH (b) σ_f of fuel type 311 (c) S_v of fuel type 311

(d) h of fuel type 311 (e) ρ_p (f) Wind speed norm

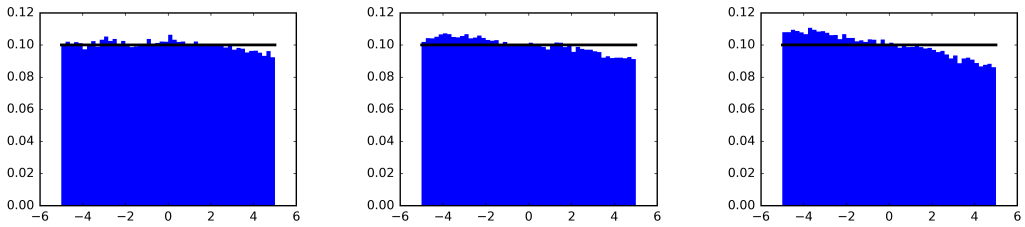
(g) m_c (h) Wind direction (i) Time of fire end

Figure 1: Histograms of some marginals of the calibrated distribution obtained with $\beta = 1/2$. The histograms are normalized to allow comparison with the PDFs of the prior distribution (bold black line). Only the perturbations whose distributions are most different from that of the priors are represented.

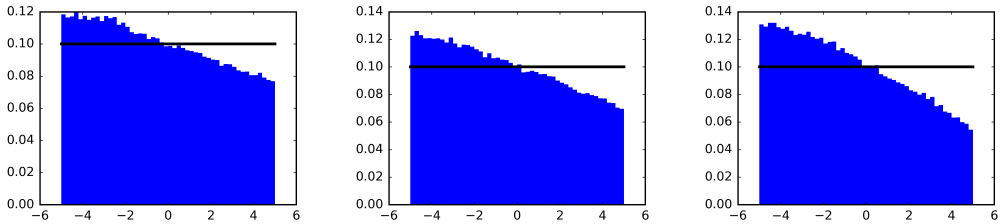
direction, m_c, and time of fire end. Similar results are obtained for other values of β, with an increasing deviation from the prior when β increases. The variables with the most deviation are the same, except for the perturbations on the fuel load σ_f of fuel type 311, which does not differ much with higher β. On the other hand, for $\beta = 1$ and $\beta = 2$, the marginals of the other individual fuel parameters S_v and h of fuel types 321 (Natural grassland) and 323

(Sclerophyllous vegetation) also show a larger deviation from those of the prior distribution, but not as much as for fuel type 311.

The variables whose marginals change the most after calibration are presumably those that have the most influence on the (emulated) energy function and therefore on the simulations. The ROS is proportional to ΔH and is highly sensitive to wind speed, and the new distributions for these inputs favor a reduced ROS, which gives more probability to smaller simulated burned areas. This tendency to favor a reduced ROS increases with β as can be seen in Figures 2 and 3. The calibrated distribution of the perturbation on the time of fire end has the same effect as it favors shorter fire durations. The maximum perturbations of the time of fire start are lower than for the time of fire end for all fire cases. This difference in uncertainty is probably the reason why the distribution on time of fire end is more affected by the calibration than is that of time of fire start, although they both have an influence on the duration of the simulated fire. Similarly, we can assume that the simulated burned surface is less sensitive to the location of the ignition point compared to the other sources of uncertainty. Individual fuel parameters influence the ROS but only in some regions of the simulation domain, which comprises two to six major fuel types depending on the fire case. For instance, fuel type 311 is involved in four of the seven fires (Chiatra, Sant'Andrea di Cotone, Nonza and Ghisoni) so its fuel parameters are more likely to be influential than these of fuel type 322 (Moors and heathland), which is almost only involved in Ghisoni fire case. For all values of β, the calibrated distribution of the perturbation of ρ_p, the particle density, favors high values. The mean of the perturbation of wind direction remains close to 0 regardless of the value of β, which is an indication of unbiased meteorological data, and its standard deviation decreases with β (see Figure 4). Interestingly, the distribution of the perturbation of the fuel moisture m_c at high values of β is dome-shaped, with a mean close to that of the prior (see Figure 5).

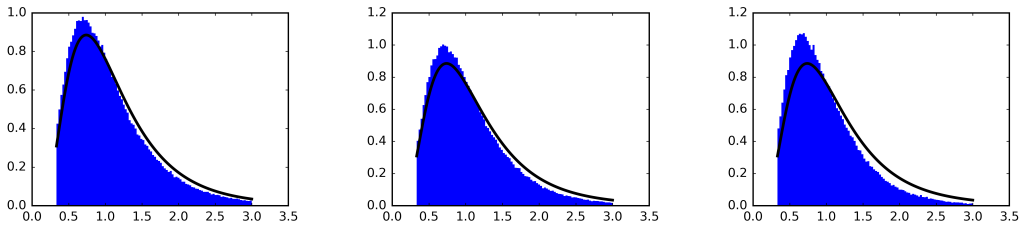


(a) $\beta = 1/10$ (b) $\beta = 1/7$ (c) $\beta = 1/4$

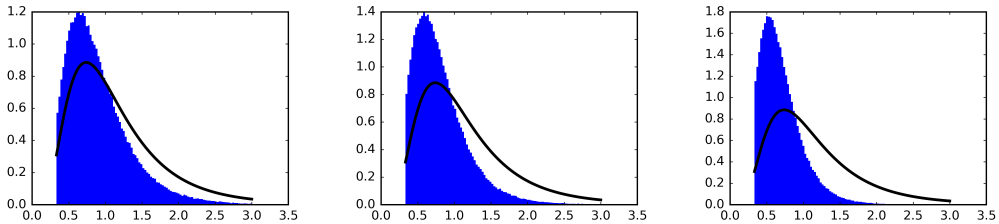


(d) $\beta = 1/2$ (e) $\beta = 1$ (f) $\beta = 2$

Figure 2: Marginal calibrated distribution of ΔH for different values of β. The bold black line indicates the PDF of the prior distribution.

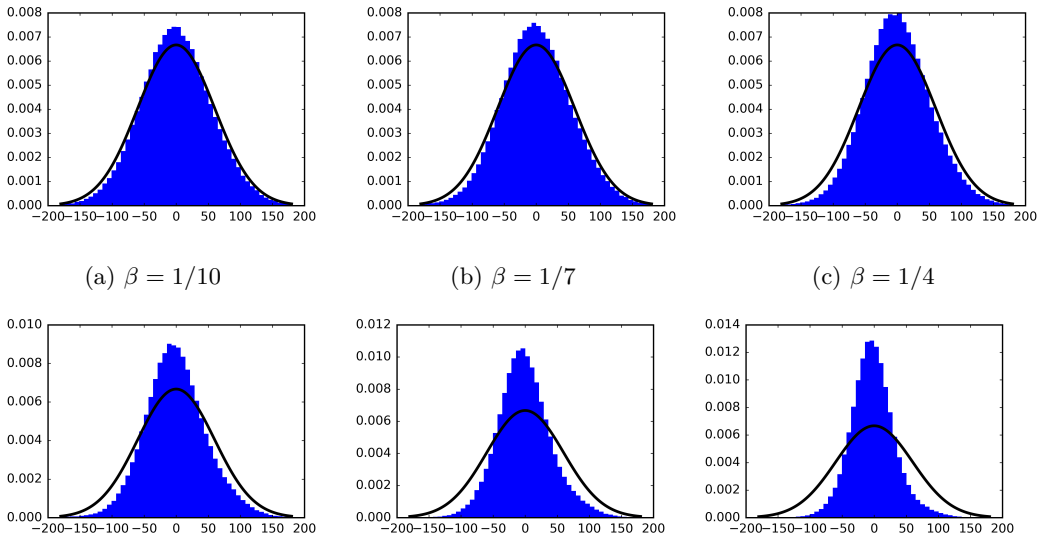


(a) $\beta = 1/10$ (b) $\beta = 1/7$ (c) $\beta = 1/4$



(d) $\beta = 1/2$ (e) $\beta = 1$ (f) $\beta = 2$

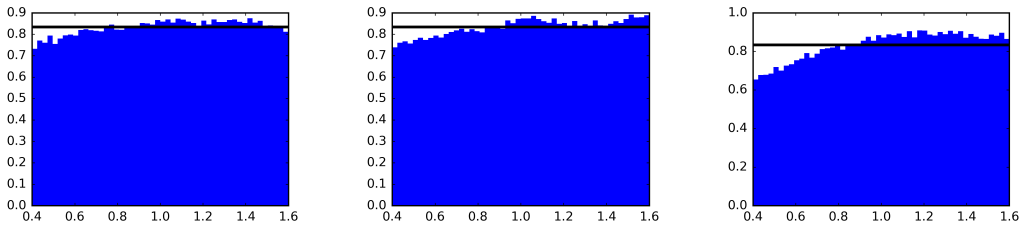
Figure 3: Marginal calibrated distribution of wind speed norm for different values of β. The bold black line indicates the PDF of the prior distribution.



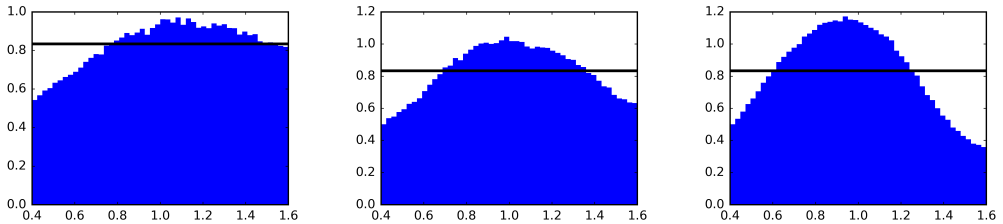
(a) $\beta = 1/10$ (b) $\beta = 1/7$ (c) $\beta = 1/4$

(d) $\beta = 1/2$ (e) $\beta = 1$ (f) $\beta = 2$

Figure 4: Marginal calibrated distribution of wind direction for different values of β. The bold black line indicates the PDF of the prior distribution.



(a) $\beta = 1/10$ (b) $\beta = 1/7$ (c) $\beta = 1/4$



(d) $\beta = 1/2$ (e) $\beta = 1$ (f) $\beta = 2$

Figure 5: Marginal calibrated distribution of the fuel moisture content m_c for different values of β. The bold black line indicates the PDF of the prior distribution.

4.3. Ensemble evaluation

The Brier skill scores of the calibrated, prior and reference ensembles for all seven fire cases are presented in Table 2. The size of the reference ensembles is 500 members for all fires other than Ghisoni (324 members) which is smaller than that of the prior and calibrated ensembles. With 10000 bootstrap samples for each ensemble, we deduced approximate 95% confidence intervals that are reported in Table 3.

Fire name (ensemble size)	Reference	Prior	$\beta = 1/10$	$\beta = 1/7$	$\beta = 1/4$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$
Calenzana (10000)	0.269	0.291	0.304	0.308	0.309	**0.314**	0.308	0.284
Chiatra (10000)	0.324	**0.386**	0.385	0.379	0.371	0.358	0.342	0.325
Ville di Paraso (2000)	0.021	0.168	0.179	0.182	**0.189**	0.188	0.176	0.168
Sant'Andrea di Cotone (5000)	0.190	0.408	0.429	0.442	0.454	0.468	0.485	**0.494**
Olmeta di Tuda (2000)	0.063	0.187	0.230	0.219	0.278	0.322	0.378	**0.451**
Nonza (4000)	-5.323	-3.089	-3.124	-3.133	-3.124	**-3.044**	-3.057	-3.053
Ghisoni (2000)	-9.986	-10.273	-9.831	-9.851	-9.333	-9.018	-8.638	**-8.332**
Global	-1.609	-1.332	-1.266	-1.269	-1.191	-1.135	-1.080	**-1.033**

Table 2: Brier skill score of the reference, prior and calibrated ensembles for the seven fire cases individually and globally (last line). For a given fire case, the best value of the BSS is represented in bold. The ensemble size applies to all ensembles except the reference ensembles that are of size 500 for all fires other than Ghisoni (324).

Fire name (ensemble size)	Reference	Prior	$\beta = 1/10$	$\beta = 1/7$	$\beta = 1/4$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$
Calenzana (10000)	[0.238, 0.300]	[0.284, 0.297]	[0.298, 0.310]	[0.301, 0.314]	[0.303, 0.315]	[0.308, 0.320]	[0.303, 0.314]	[0.278, 0.290]
Chiatra (10000)	[0.316, 0.332]	[0.383, 0.389]	[0.382, 0.389]	[0.376, 0.383]	[0.367, 0.375]	[0.354, 0.362]	[0.338, 0.346]	[0.322, 0.329]
Ville di Paraso (2000)	[-0.001, 0.042]	[0.156, 0.179]	[0.170, 0.188]	[0.173, 0.192]	[0.181, 0.197]	[0.181, 0.194]	[0.170, 0.182]	[0.162, 0.174]
Sant'Andrea di Cotone (5000)	[0.159, 0.306]	[0.398, 0.419]	[0.419, 0.438]	[0.433, 0.451]	[0.446, 0.462]	[0.461, 0.475]	[0.480, 0.491]	[0.489, 0.499]
Olmeta di Tuda (2000)	[0.011, 0.115]	[0.161, 0.214]	[0.204, 0.256]	[0.193, 0.245]	[0.254, 0.303]	[0.299, 0.346]	[0.356, 0.400]	[0.433, 0.470]
Nonza (4000)	[-5.458, -5.187]	[-3.166, -3.011]	[-10.017, -9.645]	[-3.213, -3.053]	[-3.203, -3.045]	[-3.125, -2.963]	[-3.136, -2.978]	[-8.472, -8.192]
Ghisoni (2000)	[-10.118, -9.854]	[-10.477, -10.069]	[-10.017, -9.645]	[-10.036, -9.665]	[-9.510, -9.156]	[-9.187, -8.850]	[-8.784, -8.492]	[-8.472, -8.192]
Global	[-1.635, -1.586]	[-1.359, -1.305]	[-1.291, -1.241]	[-1.294, -1.245]	[-1.215, -1.168]	[-1.158, -1.113]	[-1.100, -1.060]	[-1.053, -1.014]

Table 3: Bootstrap 95% confidence intervals of the BSS values from Table 2. The ensemble size applies to all ensembles except for the reference ensemble whose size is 500 for all fires other than Ghisoni (324)

There is considerable improvement of the BSS for almost all fire cases with the prior

ensembles compared to the reference ensembles. It is also the case with the calibrated ensembles. However, although most calibrated ensembles have better BSS than the reference ensembles, there is not always an improvement compared to the prior ensembles. For the fires of Sant'Andrea di Cotone, Olmeta di Tuda and Ghisoni, there is an overall increase of BSS with β. For the fire of Chiatra, it is the opposite: the Brier skill score decreases with β. For the fires of Calenzana and Ville di Paraso, the variation of BSS with β is relatively low, and there is probably an optimum for intermediate values of β. There might also be an optimum at intermediate β for the fire of Nonza, but the confidence intervals are too large to support this assumption.

There is not an indisputable best value β for which the accuracy is optimal for all fire cases. For $\beta = 2$, we have the best BSS for three fire cases, but for three of the remaining four cases, the BSS is in the lowest among the calibrated ensembles. Still, the global BSS is the best for $\beta = 2$, because the increase in accuracy for some fires is more significant than that of the other calibrated ensembles, while the decrease of accuracy for the other fire cases is relatively low.

In order to determine which value of β might be the most appropriate, we can instead compare the overall ranking of the Brier skill scores. These rankings are reported in Table 4. According to this method, it results that the best distribution is the one corresponding to $\beta = 1/2$.

In Figure 6 the global (i.e., "average" of the seven fires) rank histogram is represented for four distributions: reference, prior and two calibrated ($\beta = 1/4$ and $\beta = 2$). All distributions lead to higher bars on the left of the rank histogram, which is characteristic of a tendency to overpredict the burn probabilities. Calibration limits overprediction and leads to a histogram that is closer to the ideal uniform histogram.

The reliability and sharpness diagrams of these distributions are represented in Figure 7. Although the calibrated distribution lead to overall slightly more reliable ensembles, there

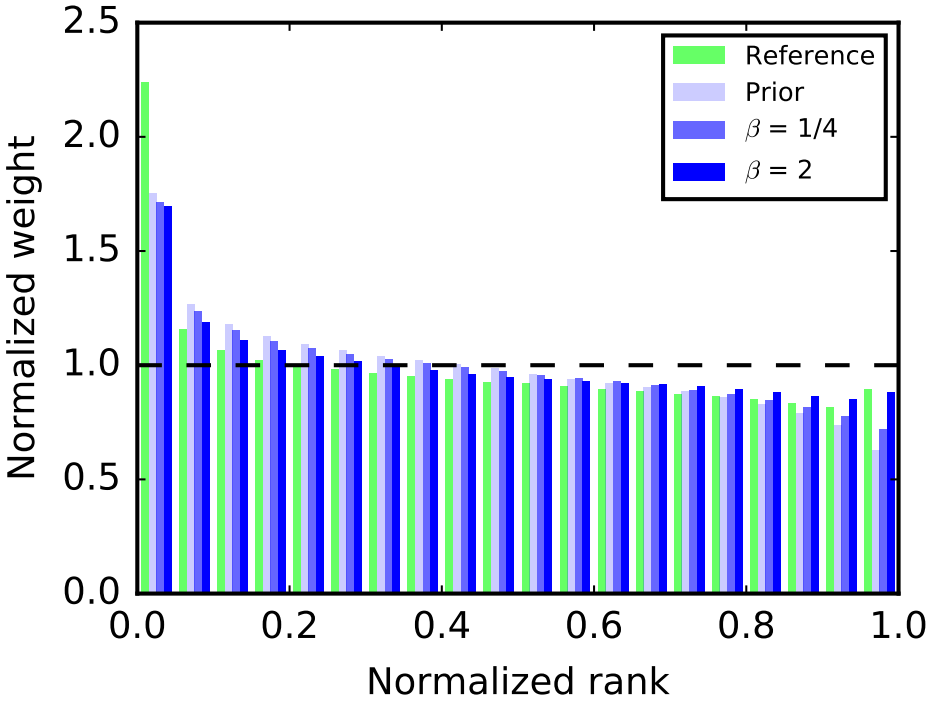


Figure 6: Global rank histogram for several ensembles. The dotted line indicates the ideal histogram of a consistent ensemble

Fire name (ensemble size)	Reference	Prior	$\beta = 1/10$	$\beta = 1/7$	$\beta = 1/4$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$
Calenzana (10000)	8	6	5	3	2	1	3	7
Chiatra (10000)	8	1	2	3	4	5	6	7
Ville di Paraso (2000)	8	6	4	3	1	2	5	6
Sant'Andrea di Cotone (5000)	8	7	6	5	4	3	2	1
Olmeta di Tuda (2000)	8	7	5	6	4	3	2	1
Nonza (4000)	8	4	5	7	5	1	3	2
Ghisoni (2000)	7	8	5	6	4	3	2	1
Sum	55	39	32	33	24	18	23	25
Overall ranking	8	7	5	6	3	1	2	4

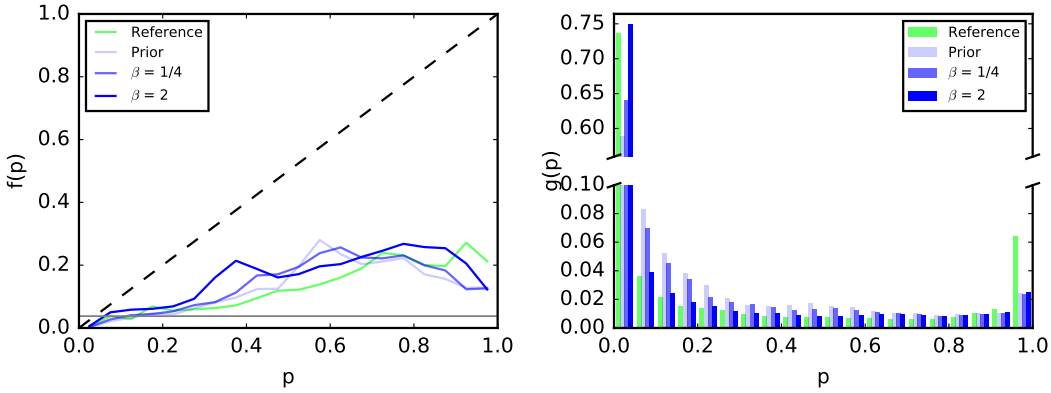
Table 4: Ranking of the Brier skill scores in Table 2. The ensemble size applies to all ensembles except the reference ensembles that are of size 500 for all fires other than Ghisoni (324)

is not an improvement for all predicted probabilities p.

The effect of calibration can also be investigated on individual fires. The maps of burn probability of the fires of Calenzana and Olmeta di Tuda are represented in Figures 8 and 9, respectively. We focus on the reference and prior ensembles as well as the calibrated ensemble corresponding $\beta = 2$. The probabilities "spread" further for the prior ensembles compared to the reference, which is due to the larger uncertainty on the input parameters. Calibration limits the spread because the calibrated probability distributions mostly favors lower rate of spread and because the uncertainty on wind direction is much lower.

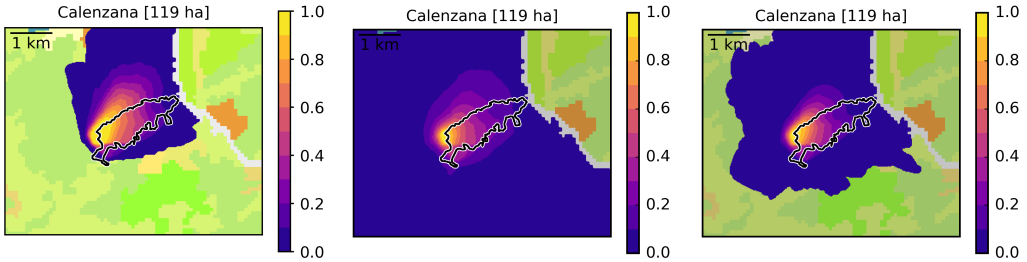
5. Discussion and conclusions

The proposed approach led to the generation of calibrated ensembles whose input distributions are defined by a posterior PDFs with a pseudo-likelihood function that involves the Wasserstein distance between simulated and observed burned surfaces of several fire cases. Due to the high dimension and the computational requirements of the pseudo-likelihood function, a Gaussian process emulator was built to obtain a sample of the calibrated input



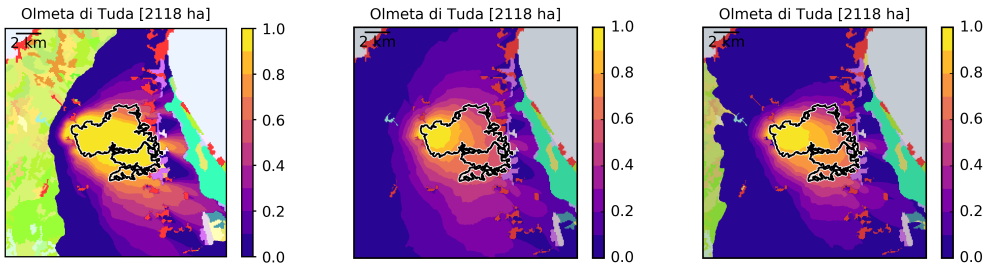
(a) Reliability. (b) Sharpness

Figure 7: Reliability and sharpness diagrams for several ensembles. (a) The dotted black line indicates a reliable prediction system and the solid gray line the optimal probability p_c of an ensemble with constant predicted probability.



(a) Reference (b) Prior (c) Calibrated, $\beta = 2$

Figure 8: Burn probability maps of Calenzana fire for several ensembles.
The colorbar indicates the predicted burn probability ; the black and white line is the contour of the observed burned surface ; background colors represent the Corine Land Cover data [24]



(a) Reference (b) Prior (c) Calibrated, $\beta = 2$

Figure 9: Same legend as Figure 8 but for the fire of Olmeta di Tuda

distribution with a MCMC algorithm in about one day of computation on 8 CPU cores.

The emulation with Gaussian process of the pseudo-likelihood of the posterior distribution shows a good accuracy ($SMSE < 5\%$), and mostly led to calibrated distributions that favored lower rate of spread and lower uncertainty on wind direction. It resulted in fire spread ensembles with overall lower fire spread and the resulting burn probabilities were less dispersed. Globally on the seven studied fires, the prior ensemble had a tendency to overpredict burn probability. This issue was less and less significant in the calibrated ensembles with increasing β, the "weight" of the pseudo-likelihood function against the prior distribution. The calibration was successful in modifying the probability distribution of the input so that the fire spread predictions have better overall accuracy.

We did not try to obtain calibrated distributions with higher values of β, but we can reasonably assume that increasing β will lead to distributions that favor even lower rate of spread. Although reducing overprediction is a desirable consequence, an adverse effect is that underprediction will become more significant for fires where the prior ensemble already underpredicts burn probability, which was the case for the fire of Chiatra where calibration led to lower accuracy. In our simulations, we did not model the firefighting actions, so overpredicting of the burn probabilities is preferable to underpredicting. Moreover, underprediction might result in an operationally non acceptable result with areas that might be

burned but not simulated by any ensemble member.

This raises the question of the choice of the pseudo-likelihood function. First, we used the parameter β to adjust its weight, and the chosen values were rather arbitrary. Clearly, low values of β will lead to a distribution that is very close to the prior. The rankings of the BSS for each values of β show a best overall for $\beta = 1/2$ that does not result in the best global BSS, but instead a very good one for most fires.

Ideally, β should be representative of the error between simulation and observation: higher error should imply lower β. Yet, our energy function for the comparison between observation and simulation relies on a weighted sum of (squared) Wasserstein distances, whose values are difficult to interpret. The Wasserstein distance remains a suitable metric for surface comparison that allowed us to increase the overall accuracy of our ensemble predictions. Hopefully, given appropriate probabilistic assumptions on the error between observation and simulation, we could derive a likelihood function that makes use of the Wasserstein distance, but modeling this error is not trivial since the model output is a burned surface. Similarly, a perspective is to take into account other sources of uncertainty in the calibration procedure, notably model error, but it is not straightforward due to the nature of model output.

Another open question is the choice of the fire cases used for calibration. Here, we selected all large fires for one season and one region, which is still a low number of fires. Even so, we used all of them for calibration and obtained an overall increase in accuracy. However, there is no guarantee that there will still be an overall improvement if we consider other fires and we might consider this calibration valid only for this season and this region. Ideally, there should be a sufficient number of fires that constitute the training basis of the calibration to ensure its application in a wide range of conditions. The evaluation of the ensembles could be carried out not only on the fires used for training but also on other fires that play the role of test sample. More fires in the training sample would provide more information, which

should limit over-fitting and potentially lead to calibrated input distributions that are more different from the prior.

Overall, we proposed a promising mathematical method to calibrate the probabilistic predictions of wildland fire spread. Improving prediction accuracy is crucial especially in the field of wildland fires where human lives, infrastructures and ecosystems are endangered. We underlined several points in the method that could be the subject of further work. Main research perspective is now to combine these calibrated ensembles with models for probability of ignition and values at stake to assess next day wildfire risk, which is relevant to fire managers, and help in the decision of firefighting actions and fire prevention planning.

Acknowledgments

This research was carried out within the ANR-16-CE04-0006 FIRECASTER.

Conficts of interest

Declarations of interest: none.

Appendix

Appendix A. Probabilistic evaluation

We consider an evaluation domain $\mathcal{X}$ that encompasses $\mathcal{S}_{obs}$, the observed burned surface, and consider a regular grid on $\mathcal{X}$ that comprises N points $x_1, ..., x_N$. Now defining $p_i = \mathbb{P}[x_i \in \mathcal{S}_{\boldsymbol{u}}]$ and $o_i = 1$ if $x_i \in \mathcal{S}_{obs}, 0$ otherwise, we define the Brier score (BS) as follows:

$$BS = \sum_{i=1}^{N} (o_i - p_i)^2. \tag{A.1}$$

This score ranges between 0 and 1 and is negatively oriented. Among the ensembles that forecast a constant probability, the one with the lowest Brier score is obtained with the probability $p_c = \frac{1}{N}\sum_{i=1}^{N} o_i$, and the Brier score of this ensemble is $BS_c = p_c(1-p_c)$. We define the Brier skill score (BSS) as follows:

$$BSS = 1 - \frac{BS}{BS_c}. \tag{A.2}$$

When there are several fires, we propose to summarize the Brier scores by their mean. The corresponding value of BS_c is obtained with the mean of the p_c. With these two global Brier scores, we define the global Brier skill score as in equation (A.2).

In practice, the probabilities p_i are estimated with a Monte Carlo method. We have an ensemble of n independently sampled input vectors $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$ and their corresponding burned surfaces $\mathcal{S}_{\boldsymbol{u}_1}, ..., \mathcal{S}_{\boldsymbol{u}_n}$. For all i, the estimate of the burn probability p_i with our finite ensemble is $\hat{p_i} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}_j})$, where $\mathbb{1}$ stands for the indicator function. The value of BS is unknown, so in practice, we estimate it with $\widehat{BS}$ which is computed following equation (A.1) by using $\hat{p_i}$ instead of p_i. We are interested in knowing how accurate the estimation $\widehat{BS}$ of BS is. We propose to estimate the standard deviation of the estimator with bootstrap [25]. For a bootstrap sample, we re-estimate the probabilities p_i by sampling with replacement among the simulated burned surfaces $\mathcal{S}_{\boldsymbol{u}_1}, ..., \mathcal{S}_{\boldsymbol{u}_n}$. With a large enough set of bootstrap samples, we obtain σ^b_{BS}, an estimator of the standard deviation of $\widehat{BS}$. Provided that the estimator of the Brier score is asymptotically normal and that the bootstrap estimation is consistent, we can approximate a confidence interval at level $1-\alpha$ for $\widehat{BS}$ by:

$$\widehat{BS} \pm z_{1-\alpha/2}\ \sigma^b_{BS}, \tag{A.3}$$

where z_q is the quantile of the standard normal distribution for probability q. The Brier skill

score being a function of BS, bootstrap can also be carried out to obtain an approximate confidence interval.

Now, we follow some aspects of bootstrap theory highlighted by Shao [26] to prove the required properties for the confidence interval. We denote by F the probability measure of the random vector $\boldsymbol{U}$. We have the following expression for the simulated burn probabilities: $p_i = \int_{\mathbb{R}^d} \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}}) dF(\boldsymbol{u})$. We denote the Brier score corresponding to the probability distribution F as $BS(F)$. First, we compute the influence function L_F for distribution F. Denoting $\delta_{\boldsymbol{u}}$ as the Dirac delta distribution at point $\boldsymbol{u}$ and setting $\epsilon > 0$, we have

$$BS((1-\epsilon)F + \epsilon\delta_{\boldsymbol{u}}) = \frac{1}{N}\sum_{i=1}^{N}(o_i - (1-\epsilon)p_i - \epsilon\mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}}))^2, \tag{A.4}$$

which yields

$$\begin{aligned}\frac{BS((1-\epsilon)F + \epsilon\delta_{\boldsymbol{u}}) - BS(F)}{\epsilon} =& \frac{1}{N}\sum_{i=1}^{N}(2o_i - (2-\epsilon)p_i + \epsilon\mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}})))(p_i - \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}})) \\ &\underset{\epsilon\to 0}{\to} \frac{1}{N}\sum_{i=1}^{N}(2o_i - 2p_i)(p_i - \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}}))) = L_F(\boldsymbol{u}).\end{aligned}$$

$L_F(\boldsymbol{u})$ is the influence function of the Brier score at point $\boldsymbol{u}$. As expected, we notice that $\int L_F(\boldsymbol{u})dF(\boldsymbol{u}) = \frac{1}{N}\sum_{i=1}^{N}(2o_i - 2p_i)(p_i - p_i) = 0$. Now, let us prove the Fréchet differentiability of the Brier score. We use the influence function of the Brier score and we define another probability measure G for which $q_i = \int \mathbb{1}(x_i \in \mathcal{S}_{\boldsymbol{u}})dG(\boldsymbol{u})$. We have

$$\left|BS(G) - BS(F) - \int L_F(\boldsymbol{u}) d[G-F](\boldsymbol{u})\right| = \left|\frac{1}{N}\sum_{i=1}^{N}(o_i - q_i)^2 - (o_i - p_i)^2 - (2o_i - 2p_i)(p_i - q_i)\right|$$
$$= \left|\frac{1}{N}\sum_{i=1}^{N}(2o_i - q_i - p_i)(p_i - q_i) - (2o_i - 2p_i)(p_i - q_i)\right|$$
$$= \frac{1}{N}\sum_{i=1}^{N}(q_i - p_i)^2$$

For two probability measures P and Q defined on the same measurable space $(\Omega, \mathcal{F})$, we define the total variation distance $||P - Q|| = \sup_{A \in \mathcal{F}} |P(A) - Q(A)|$. Since p_i and q_i are probabilities of the same event but according to the distributions F and G respectively, we have $|q_i - p_i| \leq ||G - F||$. It follows that

$$\frac{\left|BS(G) - BS(F) - \int L_F(\boldsymbol{u}) d[G-F](\boldsymbol{u})\right|}{||G-F||} \leq \frac{1}{N}\sum_{i=1}^{N}\frac{||G-F||^2}{||G-F||}$$
$$= ||G-F||$$
$$\underset{||G-F||\to 0}{\longrightarrow} 0,$$

and since $\int L_F(\boldsymbol{u}) dF(\boldsymbol{u}) = 0$, we conclude that the Brier score is Fréchet differentiable. From [26], our estimator of the Brier score is asymptotically normal with the following distribution:

$$\sqrt{n}(\widehat{BS} - BS) \underset{n\to\infty}{\longrightarrow} \mathcal{N}\left(0, \int L_F(\boldsymbol{u})^2 dF(\boldsymbol{u})\right), \tag{A.5}$$

(provided that the variance in equation (A.5) is finite) and the bootstrap estimator is consistent. All that is left is to calculate

$$\begin{aligned}\int_{R^d} L_F(\boldsymbol{u})^2 dF(\boldsymbol{u}) &= \frac{1}{N^2}\int_{R^d}\sum_{i,j=1}^{N}(2o_i-2p_i)(p_i-\mathbb{1}(x_i\in\mathcal{S}_{\boldsymbol{u}}))(2o_j-2p_j)(p_j-\mathbb{1}(x_j\in\mathcal{S}_{\boldsymbol{u}}))\\ &= \frac{4}{N^2}\sum_{i,j=1}^{N}(o_i-p_i)(o_j-p_j)(p_ip_j-p_ip_j-p_jp_i+\int_{R^d}\mathbb{1}(x_i\in\mathcal{S}_{\boldsymbol{u}}\cap x_j\in\mathcal{S}_{\boldsymbol{u}})dF(\boldsymbol{u})\\ &= \frac{4}{N^2}\sum_{i,j=1}^{N}(o_i-p_i)(o_j-p_j)(p_{ij}-p_ip_j),\end{aligned}$$

which is finite and where p_{ij} is the probability of having both locations x_i and x_j burned according to distribution F. The asymptotic normality of $\widehat{BS}$ is more explicitly given by

$$\sqrt{n}(\widehat{BS}-BS)\underset{n\to\infty}{\longrightarrow}\mathcal{N}\left(0,\frac{4}{N^2}\sum_{i,j=1}^{N}(o_i-p_i)(o_j-p_j)(p_{ij}-p_ip_j)\right). \tag{A.6}$$

Appendix B. Wasserstein distance for measuring dissimilarity between two surfaces

The Wasserstein distance is a metric that compares two measures so it is suited to the comparison of probability distributions. It can be defined in several ways and appears in the more general field of optimal transport. We refer the reader to the book [27] for a more extensive review.

Given two separable metric spaces $\mathcal{X}$ and $\mathcal{Y}$ on which are defined the measures μ and ν respectively, the optimal transport problem as formulated by Kantorovitch consists in finding the infimum

$$\inf\left\{\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\ d\gamma(x,y)\ \middle|\ \gamma\in\Gamma(\mu,\nu)\right\}, \tag{B.1}$$

where $c:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}_+\cup\{\infty\}$ is a measurable function and $\Gamma(\mu,\nu)$ is the ensemble of the measures defined on $\mathcal{X}\times\mathcal{Y}$ such that their conditional measure on $\mathcal{X}$ is μ and their

conditional measure on $\mathcal{Y}$ is ν. The optimal transport problem can be defined in different equivalent ways such as the dual formulation and the Benamou-Brenier formulation, possibly up to some factor. The original formulation by Monge, on the other hand, is a bit different and not always equivalent to (B.1). The function c can be interpreted as a cost and γ as a mapping, so that $c(x, y)$ quantifies what is required to move x to y and $\gamma(x, y)$ is the amount of mass that is moved.

The Wasserstein distance, that we denote as $\mathcal{W}_2(\mu, \nu)$, is obtained in a specific case where $\mathcal{X} = \mathcal{Y} \subset \mathbb{R}^q$ and c is the squared Euclidean distance on $\mathbb{R}^q$. It is defined as the square root of the infimum of the optimal transport problem. In other words, we have

$$\mathcal{W}_2^2(\mu, \nu) = \inf \left\{ \int_{\mathbb{R}^q \times \mathbb{R}^q} ||x - y||_2^2 \, d\gamma(x, y) \;\middle|\; \gamma \in \Gamma(\mu, \nu) \right\}. \tag{B.2}$$

Typically, $q \leq 3$, and $\mathcal{W}_2^2(\mu, \nu)$ can be interpreted as the minimum amount of energy that is required to move the mass that is distributed according to μ so that, after transport, it is distributed according to ν. In the following, when we mention the Wasserstein distance, we will be referring to its square $\mathcal{W}_2^2(\mu, \nu)$.

The Wasserstein distance can be used to compare two burned surfaces. In this case, we have $q = 2$ and, as introduced in section 2, we denote $\mathcal{S}_{obs}$ as the observed burned surface and $\mathcal{S}_{\boldsymbol{u}}$ as the simulated burned surface obtained with the perturbation vector $\boldsymbol{u}$. We consider uniform probability distributions over each surface. The PDF associated to μ is defined as follows: $\forall x \in \mathbb{R}^2, \mu(x) = \frac{1}{|\mathcal{S}_{obs}|} \mathbb{1}(x \in \mathcal{S}_{obs})$, where $|\mathcal{S}|$ is the area of the burned surface $\mathcal{S}$. Similarly for ν, we have $\forall y \in \mathbb{R}^2, \nu(y) = \frac{1}{|\mathcal{S}_{\boldsymbol{u}}|} \mathbb{1}(y \in \mathcal{S}_{\boldsymbol{u}})$. The Wasserstein distance between these two probability distributions is the metric that we propose to use in order to compare these two surfaces.

Except in some cases, there is no simple analytic formula for the Wasserstein distance. An approximation can be obtained numerically via a discretization of the probability distri-

butions. Based on an orthogonal uniform grid that covers the burned surface, we identify the points $x_1, ..., x_J$ that belong to the burned surface and the PDF can be approximated by $\frac{1}{J}\sum_{j=1}^{J}\delta_{x_j}$, where δ_{x_j} is the Dirac delta distribution at point x_j. The Wasserstein distance is then computed between two discrete probability distributions: $\hat{\mu}(x) = \frac{1}{J}\sum_{j=1}^{J}\delta_{x_j}(x)$ and $\hat{\nu}(y) = \frac{1}{K}\sum_{k=1}^{K}\delta_{y_k}(y)$ where the x_j belong to $\mathcal{S}_{obs}$ and the y_k belong to $\mathcal{S}_{\boldsymbol{u}}$. In this discrete setting, the admissible distributions γ can be represented by a matrix of size $J \times K$ where each cell γ_{jk} is positive and indicates the "probability mass" that is transferred from x_j to y_k. In this case, the infimum of (B.2) is reached and is the solution of the following linear programming problem:

$$\mathcal{W}_2^2(\hat{\mu}, \hat{\nu}) = \sum_{j=1}^{J}\sum_{k=1}^{K}\gamma_{jk}||x_j - y_k||_2^2 \tag{B.3}$$

$$\text{subject to } \gamma_{jk} \geq 0, \ \sum_j \gamma_{jk} = \frac{1}{K}, \ \text{and} \sum_k \gamma_{jk} = \frac{1}{J}. \tag{B.4}$$

which is also referred to as the *Earth Mover's distance* (EMD) [28]. It is known from graph theory that the optimal γ is a sparse matrix that has at most $J + K - 1$ non-zero cells.

For the numerical resolution, we use the package *ot* from the Python toolbox POT [29]. For the fires we study, the spatial resolution of the grid of points depends on the size of the burned surface because the computational cost increases drastically with the number of points. For the small burned surfaces, the resolution is approximately 20 m, while for the largest ones, it is approximately 80 m.

References

[1] Andrew Sullivan. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *International Journal of Wildland Fire*, 18:349–368, 07 2009. doi: 10.1071/WF06143.

[2] Andrew Sullivan. Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models. *International Journal of Wildland Fire*, 18:387–403, 07 2009. doi: 10.1071/WF06144.

[3] Andrew Sullivan. Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models. *International Journal of Wildland Fire*, 18:369–386, 07 2009. doi: 10.1071/WF06142.

[4] Mark A. Finney, Isaac C. Grenfell, Charles W. McHugh, Robert C. Seli, Diane Trethewey, Richard D. Stratton, and Stuart Brittain. A method for ensemble wildland fire simulation. *Environmental Modeling & Assessment*, 16(2):153–167, Apr 2011. ISSN 1573-2967. doi: 10.1007/s10666-010-9241-3. URL https://doi.org/10.1007/s10666-010-9241-3.

[5] Claire Miller, James Hilton, Andrew Sullivan, and Mahesh Prakash. Spark – a bushfire spread prediction tool. In Ralf Denzer, Robert M. Argent, Gerald Schimak, and Jiří Hřebíček, editors, *Environmental Software Systems. Infrastructures, Services and Applications*, pages 262–271, Cham, 2015. Springer International Publishing. ISBN 978-3-319-15994-2.

[6] Renata Pinto, Akli Benali, Ana Sá, Paulo Fernandes, P.M.M. Soares, Rita Cardoso, Ricardo Trigo, and José Pereira. Probabilistic fire spread forecast as a management tool in an operational setting. *SpringerPlus*, 5:1205, 07 2016. doi: 10.1186/s40064-016-2842-9.

[7] Akli Benali, Ana R. Ervilha, Ana C.L. Sá, Paulo M. Fernandes, Renata M.S. Pinto, Ricardo M. Trigo, and José M.C. Pereira. Deciphering the impact of uncertainty on the accuracy of large wildfire spread simulations. *Science of The Total Environment*, 569-570: 73 – 85, 2016. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2016.06.112. URL http://www.sciencedirect.com/science/article/pii/S0048969716312852.

[8] Longyan Cai, Hong He, Yu Liang, Zhiwei Wu, and Chao Huang. Analysis of the uncertainty of fuel model parameters in wildland fire modelling of a boreal forest in north-east china. *International Journal of Wildland Fire*, 28:205–215, 01 2019. doi: 10.1071/WF18083.

[9] Jean-Baptiste Filippi, Vivien Mallet, and Bahaa Nader. Representation and evaluation of wildfire propagation simulations. *International Journal of Wildland Fire*, 23:46–57, 2013. doi: 10.1071/WF12202.

[10] Thomas J. Duff, Derek M. Chong, and Kevin G. Tolhurst. Indices for the evaluation of wildfire spread simulations using contemporaneous predictions and observations of burnt area. *Environmental Modelling & Software*, 83:276 – 285, 2016. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2016.05.005. URL http://www.sciencedirect.com/science/article/pii/S1364815216301360.

[11] Francis Fujioka. A new method for the analysis of fire spread modeling errors. *International Journal of Wildland Fire - INT J WILDLAND FIRE*, 11:193–203, 01 2002. doi: 10.1071/WF02004.

[12] Thomas J. Duff, Derek M. Chong, Peter Taylor, and Kevin G. Tolhurst. Procrustes based metrics for spatial validation and calibration of two-dimensional perimeter spread models: A case study considering fire. *Agricultural and Forest Meteorology*, 160:110 –

117, 2012. ISSN 0168-1923. doi: https://doi.org/10.1016/j.agrformet.2012.03.002. URL http://www.sciencedirect.com/science/article/pii/S0168192312000950.

[13] Frédéric Allaire, Jean-Baptiste Filippi, and Vivien Mallet. Generation and evaluation of an ensemble of wildland fire simulations. *International Journal of Wildand Fire*, 2020. doi: 10.1071/WF19073.

[14] Xu Wu, Tomasz Kozlowski, Hadi Meidani, and Koroush Shirvan. Inverse uncertainty quantification using the modular bayesian approach based on gaussian process, part 1: Theory. *Nuclear Engineering and Design*, 335:339–355, Aug 2018. ISSN 0029-5493. doi: 10.1016/j.nucengdes.2018.06.004. URL http://dx.doi.org/10.1016/j.nucengdes.2018.06.004.

[15] Akli Benali, Ana C.L. Sá, Ana R. Ervilha, Ricardo M. Trigo, Paulo M. Fernandes, and José M.C. Pereira. Fire spread predictions: Sweeping uncertainty under the rug. *Science of The Total Environment*, 592:187 – 196, 2017. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2017.03.106. URL http://www.sciencedirect.com/science/article/pii/S0048969717306186.

[16] Stephen P. Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998. doi: 10.1080/10618600.1998.10474787. URL https://www.tandfonline.com/doi/abs/10.1080/10618600.1998.10474787.

[17] Olivier Roustant, David Ginsbourger, and Yves Deville. Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software, Articles*, 51(1):1–55, 2012. ISSN 1548-7660. doi: 10.18637/jss.v051.i01. URL https://www.jstatsoft.org/v051/i01.

[18] Jeong-Soo Park and Jangsun Baek. Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram. *Computers & Geosciences*, 27(1):1 – 7, 2001. ISSN 0098-3004. doi: https://doi.org/10.1016/S0098-3004(00)00016-9. URL http://www.sciencedirect.com/science/article/pii/S0098300400000169.

[19] Bertrand Iooss, Loïc Boussouf, Vincent Feuillard, and Amandine Marrel. Numerical studies of the metamodel fitting and validation processes. *International Journal On Advances in Systems and Measurements*, 3:11–21, 2010. URL https://hal.archives-ouvertes.fr/hal-00444666.

[20] C. Rasmussen and C. Williams. Gaussian process for machine learning, 01 2006.

[21] Jean-Baptiste Filippi, Frédéric Morandini, Jacques Henri Balbi, and David RC Hill. Discrete event front-tracking simulation of a physical fire-spread model. *SIMULATION*, 86(10):629–646, 2010. doi: 10.1177/0037549709343117. URL https://doi.org/10.1177/0037549709343117.

[22] Richard C. Rothermel. A mathematical model for predicting fire spread in wildland fuels. Res. Pap. INT-115. Ogden, UT: U.S. Department of Agriculture, Intermountain Forest and Range Experiment Station. 40 p, 1972.

[23] Patricia L. Andrews, Miguel G. Cruz, and Richard C. Rothermel. Examination of the wind speed limit function in the rothermel surface fire spread model. *International Journal of Wildland Fire*, 22, 2013.

[24] J. Feranec, T. Soukup, G. Hazeu, and G. Jaffrain. *European landscape dynamics: CORINE Land Cover Data.* CRC Press, Boca Raton, USA, 2016. ISBN 9781482244663.

[25] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals,

and other measures of statistical accuracy. *Statist. Sci.*, 1(1):54–75, 02 1986. doi: 10.1214/ss/1177013815. URL https://doi.org/10.1214/ss/1177013815.

[26] Jun Shao. Bootstrap estimation of the asymptotic variances of statistical functionals. *Annals of the Institute of Statistical Mathematics*, 42(4):737–752, Dec 1990. ISSN 1572-9052. doi: 10.1007/BF02481147. URL https://doi.org/10.1007/BF02481147.

[27] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser / Springer, Cham, 2015.

[28] Nicolas Bonneel, Michiel van de Panne, Sylvain Paris, and Wolfgang Heidrich. Displacement interpolation using lagrangian mass transport. *ACM Trans. Graph.*, 30(6): 158:1–158:12, December 2011. ISSN 0730-0301. doi: 10.1145/2070781.2024192. URL http://doi.acm.org/10.1145/2070781.2024192.

[29] Rémi Flamary and Nicolas Courty. Pot python optimal transport library, 2017. URL https://github.com/rflamary/POT.